CLAIMS

What is claimed is:

1   1.   A method for performing a frequent itemset operation, the method comprising the

2   steps of:

3   performing the frequent itemset operation in a plurality of phases, wherein each phase

4   is associated with combinations that have a particular number of items;

5   during at least one phase of the plurality of phases, performing the steps of

6   determining candidate combinations that are to be evaluated during the phase;

7   grouping the candidate combinations into clusters based on which items are

8   included in said candidate combinations; and

9   processing said candidate combinations, based on said clusters, to determine

10   whether the candidate combinations satisfy a frequency criteria

11   associated with said frequent itemset operation.

1   2.   The method of Claim 1 wherein the step of grouping the candidate combinations into

2   clusters includes the step of establishing an ordering for said candidate combinations by

3   sorting the candidate combinations relative to each other based on the items within each of

4   the candidate combinations.

1   3.   The method of Claim 2 wherein the step of processing the candidate combinations

2   based on the clusters includes processing the candidate combinations in a sequence based on

3   said ordering.

1   4.   The method of Claim 1 wherein the step of grouping the candidate combinations into

2   clusters includes hashing the candidate combinations into buckets based on the items that the

3   candidate combination contain.

-41-

1   5.   The method of Claim 1 wherein the step of processing the candidate combinations

2   includes generating bitmaps for the candidate combinations, and determining how many item

3   groups of an item group population include each candidate combination based on the bitmap

4   for the candidate combination.

1   6.   The method of Claim 5 wherein the step of processing the candidate combinations

2   includes, for each cluster, performing the steps of:

3   generating a bitmap for a particular combination that is a subcombination of all

4   combinations in the cluster;

5   using the bitmap for the particular combination to generate bitmaps for all

6   combinations in the cluster;

7   using the bitmap generated for each combination in the cluster to determine how

8   many item groups include the combination; and

9   after all combinations in the cluster have been processed, discarding from volatile

10   memory the bitmap for the particular combination.

1   7.   The method of Claim 1 wherein the step of processing the candidate combinations

2   includes generating bitmaps for the candidate combinations as the candidate combinations

3   are processed in a sequence, the method further comprising the steps of:

4   generating one or more intermediary bitmaps for use in generating of a bitmap for a

5   current candidate combination; and

6   after generating the bitmap for the current candidate combination, retaining in volatile

7   memory only those intermediary bitmaps that are base bitmaps of a next

8   candidate combination in said sequence; and

-42-

9   if any intermediate bitmaps are retained, then using one or more of the intermediary

10   bitmaps to generate a bitmap for the next candidate combination in said

11   sequence.

1   8.   A method for performing a frequent itemset operation, the method comprising the

2   steps of:

3   performing the frequent itemset operation in a plurality of phases, wherein each phase

4   is associated with combinations that have a particular number of items;

5   during at least one phase of the plurality of phases, performing the steps of

6   determining candidate combinations that are to be evaluated during the phase;

7   processing said candidate combinations to determine whether the candidate

8   combinations satisfy a frequency criteria associated with said frequent

9   itemset operation, wherein the step of processing the candidate

10   combinations includes generating bitmaps for the candidate

11   combinations; and

12   using an index on non-volatile memory to store a set of bitmaps that are

13   generated during said at least one phase; and

14   during a subsequent phase of said plurality of phases, performing the steps of

15   retrieving bitmaps from said index into volatile memory; and

16   using the bitmaps retrieved from said index to generate bitmaps for candidate

17   combinations of said subsequent phase.

1   9.   The method of Claim 8 wherein the step of using an index on non-volatile memory to

2   store a set of bitmaps includes using an index that uses the combination associated with a

3   bitmap as an index key for determining where within the index to place an entry for the

4   bitmap.

1 10. The method of Claim 8 wherein:

2   the at least one phase is a phase associated with N-item combinations; and

3   the set of bitmaps includes bitmaps associated with all N-item combinations that

4   satisfy the frequency criteria.

1 11. A computer-readable medium carrying one or more sequences of instructions which,

2 when executed by one or more processors, causes the one or more processors to perform the

3 method recited in Claim 1.

1 12. A computer-readable medium carrying one or more sequences of instructions which,

2 when executed by one or more processors, causes the one or more processors to perform the

3 method recited in Claim 2.

1 13. A computer-readable medium carrying one or more sequences of instructions which,

2 when executed by one or more processors, causes the one or more processors to perform the

3 method recited in Claim 3.

1 14. A computer-readable medium carrying one or more sequences of instructions which,

2 when executed by one or more processors, causes the one or more processors to perform the

3 method recited in Claim 4.

1 15. A computer-readable medium carrying one or more sequences of instructions which,

2 when executed by one or more processors, causes the one or more processors to perform the

3 method recited in Claim 5.

1 16. A computer-readable medium carrying one or more sequences of instructions which,

2 when executed by one or more processors, causes the one or more processors to perform the

3 method recited in Claim 6.

1   17.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 7.

1   18.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 8.

1   19.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 9.

1   20.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 10.